Research Plan for CSE3000 Research Project

Redefining the Single RBO Score to Achieve Greater Accuracy

Konstantin-Asen Yordanov (5529883)

April 28, 2024

Background of the Research

This research project primarily examines the properties and applications of Rank-Biased Overlap (RBO), a rank similarity measure that satisfies top-weightedness and is appropriate for comparing indefinite non-conjoint rankings [9]. Specifically, the point estimate RBO_{EXT} is the focus of this study, in which the assumption used to compute RBO_{EXT} will be modified, and its relevance for the RBO score's accuracy will be analyzed.

In the original implementation, the agreement observed at depth k (the last elements from the visible parts of the rankings S and L) is assumed to continue indefinitely across the unseen sections of the two rankings as well [9]. RBO_{EXT} is then extrapolated out from this agreement A_k.

In this context, the problem tackled in this project is an attempt to redefine RBO_{EXT} under a more relaxed assumption such that the extrapolated score is still a useful and informative criterion. More concretely, the effects of the modified theoretical/mathematical framework on the accuracy of RBO_{EXT} will be assessed in simulation, and the results will be compared against the output of the original RBO implementation.

The literature that was consulted can be categorized as follows:

- Motivation for defining and applying rank similarity measures – [5], [1], [2]
  - Cover topics such as the absence of ground truth (due to the infinite length of the rankings), the difficulty of applying precision/recall as performance metrics in all scenarios, the problem with relying on humans for relevance judgements, and the need for comparison measures
- Applications of RBO in domains where its properties are relevant and desired – [4], [7]
  - Provide examples where the use of RBO scores improves explainability and serves as a more reliable (i.e. stable) performance measure than relevance judgements
- Mathematical foundations and an elaborate theoretical framework for RBO – [6], [9]
  - Demonstrate the theoretical foundations for Rank-Biased Precision and Rank-Biased Overlap (discussing all assumptions, desired properties, and use-cases for these measures)
- Definitions of alternative rank similarity measures that apply modifications to RBO – [3], [8]
  - Propose reworked/new similarity measures with different property trade-offs and assumptions compared to RBO (examples include metricity, entropy, and the handling of ties)

Research Question

The main research question of this study is the following: **How does redefining extrapolated RBO by altering the assumption of constant agreement for items in the unseen parts of the two rankings influence the accuracy of the RBO point estimate?**

Answering this question is achievable within the 9 weeks allocated for the project as the papers defining RBO and other similarity measures inspired by RBO provide a well-established theoretical foundation, elaborating on the mathematical formulations that guarantee the desired properties.

By proposing possible reformulations of RBO_{EXT} and analyzing their performance in terms of accuracy (defined as closeness to the real RBO score), this study aims to determine whether such attempts to redefine RBO could suggest useful improvements over the original implementation and would therefore be relevant areas for future research.

To structure and guide the research, the following sub-questions were identified:

- What could serve as a measure of accuracy for the RBO scores that are computed?
- How could a window-based approach which is dependent on the evaluation depth be applied to provide a different value of the agreement used for extrapolation?
- How could regression techniques be applied to fit a function on all agreements in the visible parts of the two rankings and output a prediction of the agreement used for extrapolation?

Method

The methodology that will be applied in this study can be broken down into the following tasks:

- **Theoretical Framework:** Provide the reworked mathematical formulations for RBO_{EXT}
  - Approach 1: Using a window of the previous agreement at depth $k-1$
  - Approach 2: Fitting a function on the agreements at all depths up to point k (regression)
  - Regression Techniques Considered: linear regression, support vector regression (using linear and non-linear kernels), isotonic regression
- **Implementation:** Implement the redefined RBO_{EXT} framework in Python
  - Use the sklearn implementations for the three regression techniques (in Approach 2)
- **Evaluation Procedure:** Run 1) the original implementation, 2) the window-based formulation, and 3) the three regression-based formulations of RBO_{EXT} on simulated rankings of varying lengths/items and compare the accuracy of the computed point estimates
- **Performance Measure (Accuracy):** How close to reality are the single RBO scores when compared to the original RBO_{EXT} implementation? In which configuration do they outperform the latter (if they perform better at all)?
  - Reality, or the real RBO score, is computed by exhaustively computing RBO across the entire depth of the two rankings (which are potentially very long or infinite in-practice).
- **Variety of Simulations:** Produce simulated rankings by varying their properties and thus analyzing the scores' performance under different conditions
  - Persistence p – dictates degree of top-weightedness
  - Lengths of the visible parts s and l – provide the evaluation depth (even/uneven rankings)
  - Analyze the patterns of convergence and uncertainty at different evaluation depths (s and l)
  - Experiment with different degrees of conjointness/non-conjointness
- **Discussion of Results:** Analyze the observed performance improvements based on the new theoretical framework (if there are none, reason about why this is the case)
  - Discuss to what extent the properties for indefinite rank similarity measures remain satisfied (top-weightedness, handling non-conjoint and/or indefinite rankings)
  - Discuss the impact of different values for $p/s/l$ on the scores' accuracy (cause-and-effect)

Planning of the Research Project

If unspecified, the intended deadline for a given sub-task is **the Sunday of the respective week**.

- Week 2
  - Present the research plan during the weekly meeting (Tuesday)
  - Establish the mathematical framework for the reformulated RBO_{EXT}
  - Write the Introduction section of the research paper
- Week 3
  - Discuss compatibility with the simulation code and any noteworthy pitfalls/edge-cases when implementing RBO during the weekly meeting (Tuesday)
  - Submit ACS Assignment 1 (Wednesday)
  - Implement the window-based formulation and at least one regression-based formulation
  - Write the Formal Problem Description section of the research paper
- Week 4
  - Discuss the most appropriate representation of results in the paper and on the poster (tables, graphs, or combinations: what works best based on experience) during the meeting (Tuesday)
  - Submit ACS Assignments 2a and 2b (Wednesday)
  - Pick a date for the final presentation with supervisor, professor, and examiner (via email)
  - Integrate the remaining regression techniques (out of linear, support vector, and isotonic)
  - Gather initial results from simulations (for now, trying 1 value for p and 1-2 values for s/l is sufficient – more varied simulations will be run after the midterm presentation)
  - Write the Own Proposed Contribution section of the paper
  - Prepare a rough draft of the final poster (to-be-used during the midterm presentation)
- Week 5
  - Present the progress thus-far during the midterm presentation (Tuesday)
  - Submit ACS Assignment 3 (Friday)
  - Continue running simulations (more values for $p/s/l$, varying degrees of conjointness)
  - Write the Experimental Setup and Evaluation Process section(s) of the paper
  - Start writing the Results section of the paper
- Week 6
  - Talk about guidelines for writing the Discussion section (how much to cross-reference, in what order to discuss the observed results and RBO's desired properties for maximal readability) during the weekly meeting (Tuesday)
  - Complete the Results section of the paper (having run additional varied simulations)
  - Fill in the tables containing the results with the newly-aggregated RBO scores
  - Write the Responsible Research section of the paper
  - Write tentative versions of the Discussion and Conclusion sections of the report (in preparation for submitting the first draft)

- Week 7
  - Submit first draft of the paper for feedback from supervisor and peers (Monday)
  - Submit peer review of another student's first draft (Wednesday)
  - Incorporate received feedback to complete the second draft of the paper – in preparation for the professor's feedback (focus on depth of analysis and logical coherence of the text)
- Week 8
  - Submit second draft of the paper for feedback from professor (Tuesday)
  - Start preparing the poster to-be-used during the final presentation (focus on intuitive layout and visual clarity – include only the study's most important steps/findings/conclusions)
- Week 9
  - Incorporate the professor's feedback to complete the final version of the research paper
  - Finalize the poster (having discussed best practices during the weekly meeting)
  - Submit the final version of the research paper (Sunday)
- Week 10
  - Submit the poster (Monday)
  - Present the study's steps/findings/conclusions during the final presentation

References

[1] Judit Bar-Ilan. "Comparing rankings of search results on the Web". In: *Information Processing & Management* 41.6 (2005). Special Issue on Infometrics, pp. 1511–1519. ISSN: 0306-4573. DOI: https://doi.org/10.1016/j.ipm.2005.03.008. URL: https://www.sciencedirect.com/science/article/pii/S0306457305000312.

[2] Judit Bar-Ilan, Mazlita Mat-Hassan, and Mark Levene. "Methods for comparing rankings of search engine results". In: *Computer Networks* 50.10 (2006). I. Web Dynamics II. Algorithms for Distributed Systems, pp. 1448–1463. ISSN: 1389-1286. DOI: https://doi.org/10.1016/j.comnet.2005.10.020. URL: https://www.sciencedirect.com/science/article/pii/S1389128605003646.

[3] Bruno Cardoso and João Magalhães. "Google, bing and a new perspective on ranking similarity". In: *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. CIKM '11. Glasgow, Scotland, UK: Association for Computing Machinery, 2011, pp. 1933–1936. ISBN: 9781450307178. DOI: 10.1145/2063576.2063858. URL: https://doi.org/10.1145/2063576.2063858.

[4] Charles L. A. Clarke, Mark D. Smucker, and Alexandra Vtyurina. "Offline Evaluation by Maximum Similarity to an Ideal Ranking". In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. CIKM '20. Virtual Event, Ireland: Association for Computing Machinery, 2020, pp. 225–234. ISBN: 9781450368599. DOI: 10.1145/3340531.3411915. URL: https://doi.org/10.1145/3340531.3411915.

[5] Ronald Fagin, Ravi Kumar, and D. Sivakumar. "Comparing Top k Lists". In: *SIAM Journal on Discrete Mathematics* 17.1 (2003), pp. 134–160. DOI: 10.1137/S0895480102412856. eprint: https://doi.org/10.1137/S0895480102412856. URL: https://doi.org/10.1137/S0895480102412856.

[6] Alistair Moffat and Justin Zobel. "Rank-biased precision for measurement of retrieval effectiveness". In: *ACM Trans. Inf. Syst.* 27.1 (Dec. 2008). ISSN: 1046-8188. DOI: 10.1145/1416950.1416952. URL: https://doi.org/10.1145/1416950.1416952.

[7] Alessia Sarica, Andrea Quattrone, and Aldo Quattrone. "Introducing the Rank-Biased Overlap as Similarity Measure for Feature Importance in Explainable Machine Learning: A Case Study on Parkinson's Disease". In: *Brain Informatics*. Ed. by Mufti Mahmud et al. Cham: Springer International Publishing, 2022, pp. 129–139. ISBN: 978-3-031-15037-1.

[8] Luchen Tan and Charles L. A. Clarke. "A Family of Rank Similarity Measures Based on Maximized Effectiveness Difference". In: *IEEE Transactions on Knowledge and Data Engineering* 27 (2014), pp. 2865–2877. DOI: 10.1109/TKDE.2015.2448541.

[9] William Webber, Alistair Moffat, and Justin Zobel. "A similarity measure for indefinite rankings". In: *ACM Trans. Inf. Syst.* 28.4 (Nov. 2010). ISSN: 1046-8188. DOI: 10.1145/1852102.1852106. URL: https://doi.org/10.1145/1852102.1852106.